

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-215180

(P2002-215180A)

(43) 公開日 平成14年7月31日 (2002.7.31)

(51) Int.Cl. ⁷	識別記号	F I	テ-マ-ト* (参考)
G 1 0 L 15/00		G 0 6 T 13/00	C 5 B 0 5 0
G 0 6 T 13/00		H 0 4 M 1/00	R 5 D 0 1 5
G 1 0 L 15/14		G 1 0 L 3/00	5 5 1 G 5 K 0 2 7
11/00			5 3 5 Z
H 0 4 M 1/00			5 5 1 A

審査請求 未請求 請求項の数 6 O L (全 5 頁) 最終頁に続く

(21) 出願番号 特願2001-9572(P2001-9572)

(22) 出願日 平成13年1月17日 (2001.1.17)

(71) 出願人 397055528

株式会社デジタル・メディア・ラボ
東京都港区赤坂6丁目6番20号

(72) 発明者 大海 啓一

東京都港区赤坂6丁目6番20号 テルウェ
ル赤坂ビル 株式会社デジタル・メディ
ア・ラボ内

(74) 代理人 100083183

弁理士 西 良久

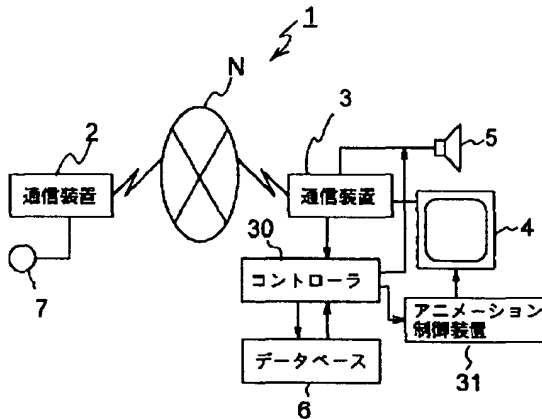
Fターム(参考) 5B050 AA08 BA07 BA08 BA12 CA05
CA08 EA24 FA02 FA10
5D015 GG03 HH23 KK02
5K027 AA11 BB01 DD14 FF01 FF22

(54) 【発明の名称】 通信装置

(57) 【要約】

【課題】 この発明は顔面アニメーション生成方法を用いて、通信装置間で音声データに同期した仮想的な顔表情の動画を滑らかに表現する通信装置に関する。

【解決手段】 受信側の通信装置に、一定の音声パターンと、これに対応し基本骨格モデルの顔表情の推移パターンとからなるデータを関連づけた動作推論データベースを設けておき、入力した音声データを分析して特徴を抽出して音声パターンを生成し、この得られた音声パターンを基に、上記動作推論データベースから顔表情の推移パターンを抽出し、上記入力した音声データと、上記顔表情の推移パターンとを関連づけると共に、受信側の通信装置に登録してある顔形状モデルに、上記顔表情の推移データから得られた顔表情の基本骨格モデルの動作をマッピングして、前記音声データに同期した仮想的な顔表情を滑らかに表現することを特徴とする。



【特許請求の範囲】

【請求項1】 通信ネットワークを介して接続される通信装置間で、送受信される音声信号を基に、受信側の通信装置のディスプレイ装置に表示された顔形状モデルの画像を音声信号に同期して動かす通信装置において、受信側の通信装置に、音声データを基にした音声パターンと、音声パターンに対応する顔表情データを基にした推移パターンとを関連づけた動作推論データベースを呼び出し可能に設けておき、
10 入力した音声信号を分析部で分析して特徴を抽出した音声パターンを生成し、
この得られた音声パターンを基に、上記動作推論データベースから顔表情の推移パターンを抽出し、
前記音声信号と、上記顔表情の推移パターンとを関連づけると共に、
受信側の通信装置に登録してある顔形状モデルに、上記顔表情の推移データから得られた推移パターンの動作をマッピングして動画像を生成し、
表示装置で前記音声信号の出力と同期して前記顔形状モデルの顔表情の動画像を表示してなることを特徴とする通信装置。

【請求項2】 顔形状モデルのデータが、予め受信側の通信装置のメモリに格納されていることを特徴とする請求項1に記載の通信装置。

【請求項3】 顔形状モデルのデータを、音声信号と共に受信側の通信装置に送信して受信側の通信装置のメモリに格納されることを特徴とする請求項1に記載の通信装置。

【請求項4】 通信装置の送受信側的一方または双方が携帯電話からなっていることを特徴とする請求項1に記載の通信装置。

【請求項5】 通信装置の送受信側的一方または双方がパーソナルコンピュータからなっていることを特徴とする請求項1に記載の通信装置。

【請求項6】 通信ネットワークが、携帯電話通信網やインターネット等のコンピュータ通信ネットワークからなっていることを特徴とする請求項1に記載の通信装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、携帯電話やインターネット等の通信ネットワーク、あるいはその組み合わせを介して接続される通信装置間で、送受信される音声信号を基に、受信側の通信装置のディスプレイ装置上に表示される顔形状モデルの動画像を音声信号に同期して動かす通信装置に関する。

【0002】

【従来の技術】従来の音声認識では音声信号の音素片を抽出し、各音素片からその時点での表情にマッピングし、それらを補間することによって連続する口の動作が

生成される。しかし、従来のこの手法では音素片を抽出した時点で音に関する表情要素（音の表情への大きさ、唇関節への音のフレージング予測等）が失われ不自然な表情動作となる。そこで、特開2000-123192号の顔面アニメーション生成方法では、現実的な顔のアニメーションを音声から直接生成する方法として、人が話をしている間に現れる顔及び顔つきの観察記録を基に、顔及び音声の動作の動的なモデルを学習し、音声を学習した顔つきのカテゴリーに関連して分析し、顔の音素間多重結合を含む長期の依存関係を無条件にモデル化する動的モデル学習ステップと、新たな入力オーディオに伴う顔つきのシーケンスである顔面シーケンスを生成する顔面シーケンス生成ステップと、前記顔面シーケンスを、顔面の制御パラメータに基づいて最適なシーケンスに変換するシーケンス変換ステップと、前記顔面の制御パラメータに基づき顔面のアニメーションを動かす顔面アニメーション動作ステップとを備え、顔面の滑らかな動画像を得る構成が示されている。

【0003】

【発明が解決しようとする課題】この発明は上記事情に鑑みて鋭意研究の結果創案されたものであって、その主たる課題は、上記顔面アニメーション生成方法を用いて、通信装置間で音声データに同期した仮想的な顔表情の動画像を滑らかに表現する通信装置を提供することにある。

【0004】

【課題を解決するための手段】前記課題を達成するために、請求項1の発明では、通信ネットワークを介して接続される通信装置間で、送受信される音声信号を基に、受信側の通信装置のディスプレイ装置に表示された顔形状モデルの画像を音声信号に同期して動かす通信装置において、受信側の通信装置に、音声データを基にした音声パターンと、音声パターンに対応する顔表情データを基にした推移パターンとを関連づけた動作推論データベースを呼び出し可能に設けておき、入力した音声信号を分析部で分析して特徴を抽出した音声パターンを生成し、この得られた音声パターンを基に、上記動作推論データベースから顔表情の推移パターンを抽出し、前記音声信号と、上記顔表情の推移パターンとを関連づけると共に、受信側の通信装置に登録してある顔形状モデルに、上記顔表情の推移データから得られた推移パターンの動作をマッピングして動画像を生成し、表示装置で前記音声信号の出力と同期して前記顔形状モデルの顔表情の動画像を表示してなる、という技術的手段を講じている。

【0005】また、請求項2の発明では、上記請求項1の発明における前記顔形状モデルのデータが、予め受信側の通信装置のメモリに格納されている、という技術的手段を講じている。請求項3の発明では、上記請求項1の発明における前記顔形状モデルのデータを、音声信号

と共に受信側の通信装置に送信して受信側の通信装置のメモリに格納される、という技術的手段を講じている。また、請求項4の発明では、上記請求項1の発明における前記通信装置の送受信側的一方または双方が携帯電話からなっている、という技術的手段を講じている。更に、請求項5の発明では、上記請求項1の発明における前記通信装置の送受信側的一方または双方がパーソナルコンピュータからなっている、という技術的手段を講じている。また、請求項6の発明では、上記請求項1の発明における通信ネットワークが、インターネット等のコンピュータ通信ネットワークからなっている、という技術的手段を講じている。

【0006】

【発明の実施の形態】以下に、この発明の通信システムおよびそのシステムで稼働するプログラムの好適実施例について図面を参照しながら説明する。図1に示す通信システム1は、送信側の通信装置2に携帯電話を用い、通信ネットワークNを介して接続される受信側の通信装置3としてパーソナルコンピュータを用いた例を示す。

【0007】受信側の通信装置3には制御部としてのコントローラ30と、外部表示装置としてのディスプレイ装置4と、スピーカー5とが接続されている。そして、通信装置3のメモリ（図示せず）には、比較的長く続く音素間多重結合を含む一定の音声データをHMM（隠れマルコフ・モデル）符号化した音声パターンと、これに対応し基本骨格モデルの顔表情データをHMM（隠れマルコフ・モデル）符号化した推移パターンとを関連づけて登録した動作推論データベース6が格納されている。

【0008】なお、コントローラ30は、アニメーション制御装置31を介してディスプレイ装置4の動画を制御するようになっている。また、図中、符号7は、送信用の音声信号を入力するマイクなどの音声入力部である。

【0009】この動作推論データベース6は、特開2000-123192号（米国特許出願番号第08/994,533号）で説明されている公知の手順による顔面アニメーション生成方法に基づき作成される。

【0010】ここでは音素片による音声認識を行うのではなく、音声データと発声に伴う表情データの変化のパターンを同時に符号化（隠れマルコフ・モデル符号化）して音声に関する状態シーケンスと表情に関する状態のシーケンスを設定し、データベース化するものであり、ある状態の音（声）のタイミングに関する顔表情の関係推移を推論するものである（図2参照）。

【0011】コントローラ30は、通信ネットワークNを介して受信側の通信装置3に入力された音声信号を、所定のサンプリング周期でサンプリングして音声データとなし、これをオーディオ分析部で分析して一定の長さの音声パターンとする。即ち、このオーディオ分析部

は、線形予測解析を用いた音声データの特徴抽出プロセスであり、図3に示すように、入力された音声データ（WAVファイル）をハニング窓を通してDFTを行いスペクトルの変換を行う。その後、対数変換、IDFTを行い、ケプストラム窓を通してスペクトルの大域的長を抽出し音声データの特徴データとする。

【0012】ここで前記受信側の通信装置3に入力された音声信号は、直接に送信されるものでも、あるいは音声メールとして送信されるものでもよい。この音声パターンは、ケプストラム解析を行ってノイズに埋もれた音声の特徴を取り出されたものである。

【0013】次いで、特徴抽出された新規な音声パターンに対して、その音声のタイミングにおける顔表情の動作の推移を予め登録してある前記動作推論データベース6から検索する。

【0014】そして、この抽出された上記顔表情の推移パターンと、上記入力された音声信号とをタイミングが一致するように関連づける。

【0015】ここで、受信側の通信装置3のメモリには、キャラクターまたは人の仮想肖像画像の3次元的な顔形状モデルデータが予め格納されている。この顔形状モデルデータは、1種類であっても、あるいは複数種類であってもよく、後者の場合は1種類が選択可能となっている。この場合、例えば、送信側の通信装置2の電話番号を予め受信側の通信装置3のメモリに登録しておき、入力した発信者の電話番号を識別データとして、それを基に顔形状モデルを選択して表示するようにしてもよい。

【0016】その他、識別データは、送信側の通信装置2から音声信号と共に送信するものであればよい。更に、送信側の通信装置2では、顔形状モデルデータを音声信号と共に受信側の通信装置3へ送信するものでもよい。そして、コントローラ30は、送信側の通信装置2より送信された音声信号が受信側の通信装置3に入力されることによって顔形状モデルを呼び出し、アニメーション制御装置31を作動させてディスプレイ装置4上に表示する。

【0017】次に、コントローラ30はアニメーション制御装置31で、この顔形状モデルのデータに、上記顔表情の推移データから得られた顔表情の基本骨格モデルの動作をマッピングさせ、運動計算を施して顔表情を動かす動画のアニメーションデータを作成する。そして、音声信号に同期するように前記ディスプレイ装置4上の顔形状モデルの顔表情を動かしてリアルな顔表情の動作を再現することができる（図4参照）。

【0018】このアニメーション制御装置31の処理は、音声信号の入力とリアルタイムで処理され、新規なアニメータを生成する。このようにして音声信号の入力に伴って、受信側の通信装置3ではディスプレイ4装置上に表示された仮想肖像画像の3次元的な顔形状の動画

5

像を滑らかに表現することができる。

【0019】この発明は、前記実施例に限定されるものではなく、図5に示すように、受信側の通信装置3がサーバであって、該サーバには動作推論データベース6が呼び出し可能に接続された構成でも良い。そして、上記サーバのWebページ(図示せず)が表示装置に対応し、このWebページ上に、予め登録された、あるいは送信側から送信された顔形状モデルを表示し、入力された音声信号を基に前記動作推論データベースから顔表情の推移パターンを前記顔形状モデルにマッピングしてアニメーションを作成し、音声信号の出力と同期させて顔表情の動画像を滑らかに表現するようにしてもよい。

【0020】同様に、送受信側の通信装置が共に携帯電話やPDA等の携帯端末であってもよく、その場合に、少なくとも受信側の携帯電話に前記受信側の通信装置と同様の構造を備えておけばよい。その他、要するにこの発明の要旨を変更しない範囲で種々設計変更しうることを勿論である。

【0021】

【発明の効果】この発明は、上記構成からなっているの
で、携帯電話やインターネット等の通信ネットワーク、
あるいはその組み合わせを介して接続される通信装置間
で、送受信される音声信号を基に、受信側の通信装置の
ディスプレイ装置上に表示される顔形状モデルの動画像

6

を音声信号に同期してリアルに動かすことができ、通話時における受信者の対話を容易にすると共に、顔形状モデルの表情の変化から発信者の感情も視覚で伝えうる通信装置サービスを提供することができ、極めて有益である。

【図面の簡単な説明】

【図1】この発明の通信装置の実施例のブロック図である。

【図2】動作推論データベースのブロック図である。

10 【図3】オーディオ分析部の機能ブロック図である。

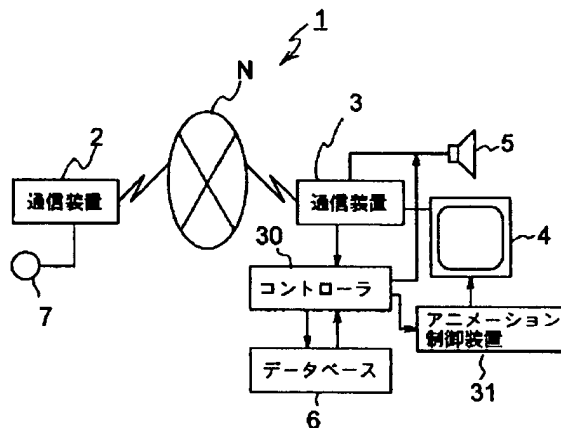
【図4】通信装置の機能ブロック図である。

【図5】通信装置の異なる実施例のブロック図である。

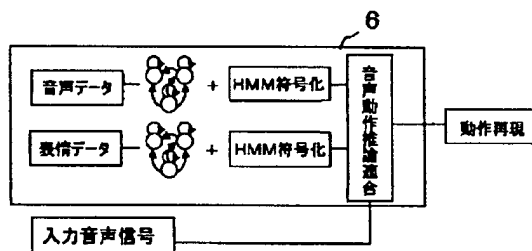
【符号の説明】

- | | |
|----|-------------|
| 1 | 通信システム |
| 2 | 送信側の通信装置 |
| 3 | 受信側の通信装置 |
| 4 | ディスプレイ装置 |
| 5 | スピーカー |
| 6 | 動作推論データベース |
| 7 | 音声入力部 |
| 30 | コントローラ |
| 31 | アニメーション制御装置 |
| N | 通信ネットワーク |

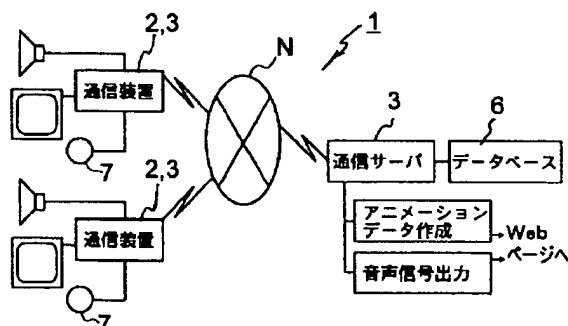
【図1】



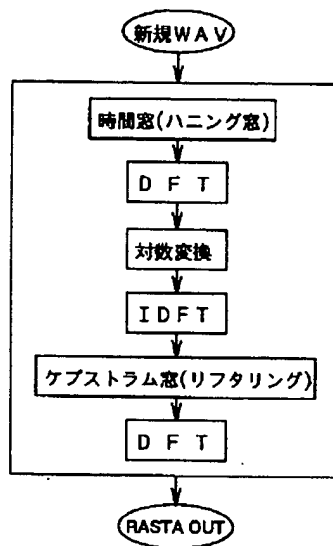
【図2】



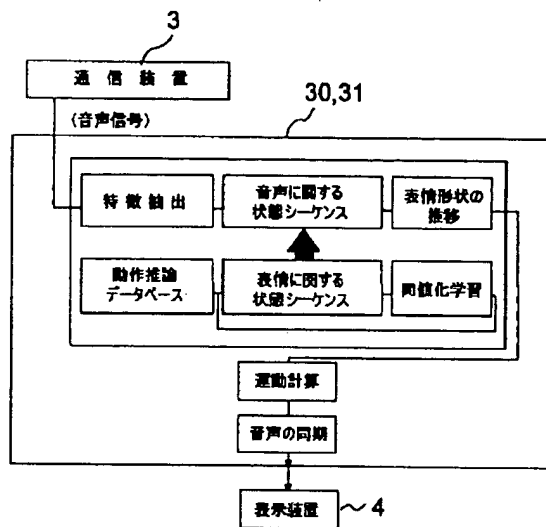
【図5】



【図3】



【図4】



フロントページの続き

(51)Int.Cl.⁷

識別記号

F I

G 1 0 L 9/16

テーマコード(参考)

CLIPPEDIMAGE= JP02002215180A

PAT-NO: JP02002215180A

DOCUMENT-IDENTIFIER: JP 2002215180 A

TITLE: COMMUNICATION DEVICE

PUBN-DATE: July 31, 2002

INVENTOR-INFORMATION:

NAME

OMI, KEIICHI

COUNTRY

N/A

ASSIGNEE-INFORMATION:

NAME

DIGITAL MEDIA LAB INC

COUNTRY

N/A

APPL-NO: JP2001009572

APPL-DATE: January 17, 2001

INT-CL (IPC): G10L015/00;G06T013/00 ;G10L015/14 ;G10L011/00
;H04M001/00

ABSTRACT:

PROBLEM TO BE SOLVED: To provide a communication device in which an animation image of an imaginary facial expression synchronized to voice data between communication devices is smoothly expressed the employing a facial animation generating method.

SOLUTION: An action reasoning database is provided in a communication device of a receiving side. The database relates a certain voice pattern and the data which are corresponded to the pattern and made up with a transition pattern of the facial expression of a basic skeleton model. Inputted voice data are analyzed, features are extracted and a voice pattern is generated. Based on the obtained voice pattern, the transition pattern of the

facial expression is
extracted from the action reasoning database and the
inputted voice data and
the transition pattern of the facial expression are
related. Then, the action
of the basic skeleton model of the facial expression
obtained from the
transition data of the facial expression is mapped onto the
facial shape model
registered in the communication device of the receiving
side and the imaginary
facial expression synchronized to the voice data is
smoothly expressed.

COPYRIGHT: (C) 2002, JPO